Substation Net Load Operational Forecasts for Fast Demand Response Implementation

Final Task Report
submitted to the
CPUC California Solar Initiative RD&D Program

January 27, 2015

Authors: Amanpreet Kaur, Lukas Nonnenmacher, Carlos Coimbra, Center for Renewable Resources and Integration, University of California, San Diego

Abstract: (Net) load forecasts for five feeders have been produced and validated for 15 min to 5 days forecast horizon. Forecasting methods like Artificial Neural Networks optimized using Genetic Algorithms, Support Vector Regression, k-Nearest Neighbors, and various other state-space and time-series models were implemented and tested successfully. The forecasting errors were found to increase on cloudy days and to increase on feeders with higher solar penetration. Exogenous inputs like day-ahead solar forecasts from the NAM model were also used as an input to further refine the forecast. The accuracy of the load forecast is limited by the NAM forecast accuracy.

UC San Diego

Subtask 5.2 Substation Net Load Operational Forecasts for Fast Demand Response Implementation

Background and Motivation

Variable rooftop solar generation contributes substantially to the power demand forecast uncertainty at substation level. The impacts of high penetration PV on distribution circuit loads such as peak demand, dynamic loading, and fast demand response capability were quantified. High-fidelity intra-hour (5, 10, 15, 30 and 60 minutes) and multiple days-ahead (1-5 days ahead) power demand forecasts methodologies for five feeders were implemented and validated.

Data and Forecast implementation

Intra-hour and day-ahead forecasts with a forecast horizon of 15 min up to 5 days-ahead were implemented for five feeders with high solar penetration i.e. Valley Center, Avocado, Creelman, Cabrillo and Alpine located in the San Diego Gas & Electric (SDG&E) operating region. The details about the five feeders are shown in Table 1. The impact of solar penetration on the load profile for the feeder with the highest solar penetration is shown Fig. 1. The variability in solar power production propagates into the load profile, which makes load forecasting more challenging during the daytime. Similar characteristics were observed in other feeder's load profile with decreasing effect of solar variability on feeder load demand with decreasing solar penetration. Solar penetration is defined as the percentage of total solar power produced w.r.t. total load on the feeder i.e., the sum of solar energy produced and net load annually.

Table 1: Details about five SDG&E feeders for which net load forecasting was implemented and tested.

SDGE feeders	Feeder length [km]	No. of customers	No. of PV systems	Solar penetration
Valley Center	51.5	471	19	23.8%
Avocado	177.8	2246	29	13.3%
Creelman	115.7	1169	43	9.26%
Cabrillo	39.6	3761	91	5.79%
Alpine	34.5	1466	28	2.39%

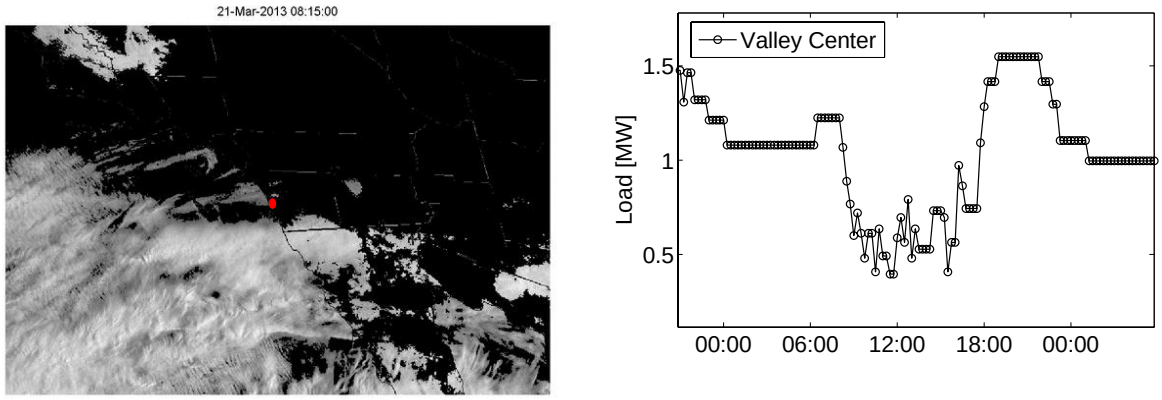

Figure 1: Sample satellite image and load profile for 21 March 2013 for a highly cloudy day. The red circle in the satellite image shows the feeder location. The effect of solar penetration on the load profile manifests in the load fluctuations during daytime [7:00 – 18:00 PDT].

For developing and testing the models, the dataset was divided into two disjoint datasets: the training set consists of data between January 2011 to June 2012 and the validation/testing set ranges from July 2012 to June 2013. For intra-hour forecasts, the methods with non-exogenous inputs were implemented, whereas for multiple day ahead forecast, day- ahead solar irradiance forecasts provided by the NOAA North American Model (NAM) were also used as inputs.

The forecast models with non-exogenous input were developed as $\dot{y}(t+k|t) = f\,(y(t), y(t-1), y(t-2) \ldots y\,(t-nk))$, where '$y$' is the net load, '$\dot{y}$' is the net load forecast issued at time 't' for 'k' steps ahead (forecast horizon) using a non-linear/linear function, 'f' with 'nk' as the number of time-lagged inputs. If no information about the underlying processes affecting the system is available, machine-learning tools like Artificial Neural Networks (ANN) and Support Vector Regression (SVR) have proven to be useful for input/output mapping. ANNs consist of multiple layers with processing units called neurons. Neurons take in weighted sum of inputs through various layers and produce an output using an activation function. For this study, the optimal structure of the ANN was obtained by using a Genetic Algorithm (GA). GA is an optimization algorithm based on survival of the fitted individuals [1]. Secondly, a non-linear regression technique known as Support Vector Regression (SVR) was applied [2]. The inputs used are known as support vectors and the function 'f' is defined as $f(y) = < w, y > + b$, where w represents the weights and b is a constant. The optimization problem is solved with the following objective and constraints,

$$\min_{w,b,\xi,\xi^*} \frac{1}{2} w'w + C \sum_{i=1}^{l} (\xi_i + \xi_i^*)$$

$$\text{subject to } y_i - (w'f(y) + b) \leq \epsilon + \xi_i,$$
$$(w'f(y) + b) - y_i \leq \epsilon + \xi_i^*,$$
$$\xi_i\ , \xi_i^* \geq 0, i = 1, \ldots, l,$$

where, $y_i \in \{y(t), y(t-1) \ldots y(t-nk)\}$ is the input; $\xi_i\ , \xi_i^*$ are the upper and lower training errors subject to an ϵ-insensitive tube and C is the cost of the error . Therefore, the parameters C, ϵ and the mapping function f control the regression quality obtained using SVR. A radial basis function was used as a mapping function while cross-validation was used to compute the other parameters [3]. Moreover, pattern recognition-based machine learning techniques i.e., k-Nearest Neighbors (kNN) were also implemented for day ahead load forecasting. This technique identifies similar patterns in the historical dataset w.r.t. the present conditions and chooses the closest neighbors. For this study 15 best neighbors were chosen. The corresponding outputs of the best neighbors are combined to issue a forecast [4]. Finally, NAM weather forecasts are used as an

input to day-ahead forecasts by utilizing an ensemble re-forecast method [5]. The ensemble re-forecast method is a combination of various linear and non-linear time-series regression techniques like autoregressive model with exogenous input (ARX), non-linear autoregressive model with exogenous input (NARX), Box-Jenkins model (BJ), etc.

Error Metrics

In this work we used the following error metrics: Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2). The MAPE measures the accuracy of a method in terms of percentage error. It is defined as MAPE [%] = $\frac{100}{n}\sum_{t=1}^{n}|\frac{A_t-F_t}{A_t}|$, where A_t is the actual value and F_t is the forecast value of the net load at time t. RMSE gives the information about the spread i.e., standard deviation in the error. A MAPE or RMSE of zero imply perfect forecasts. The coefficient of determination R^2 measures the level of dispersion about the 1:1 line in a scatter plot of measured vs. forecasted values.

Results and discussions

Intra-hour forecasts

Machine learning based forecasting models were developed for intra-hour forecasting for 15 min, and 30 min forecast horizon with 15 min averaged time-series as provided by SDG&E. Since the load profile is inherently different during daytime compared to night time due to variability introduced by solar energy, two local SVR models for each feeder were developed: one for day time and another one for night time.

Comparing the results using SVR and ANN from Table 2 and Table 3, it can be observed that the performance of both methods is similar for the 15 min forecast horizon. ANN-GA outperforms SVR for 30 min forecast horizon in terms MAPE. The sample results using local SVR are shown in Figure 2 for the Valley Center feeder to compare the accuracy of the forecasting model for clear and cloudy conditions. Most of the error occurs during the daytime especially during cloudy days. The impact of intermittency in solar power can be observed in net load, which are challenging to forecast resulting in high magnitude forecast errors. Also, despite the fact that Creelman and Alpine have lower solar penetration as compared to Valley Center and Avocado, they still have high forecasting error. These high magnitude errors for Creelman and Alpine can be attributed to the high variability in load from these feeders independent of solar generation (see Figure 3). Next, the intra-hour SVR forecast models were extended for multiple day(s) ahead forecasting.

Table 2: Net load forecasting results for 15 min forecast horizon for the five feeders.

Models	SVR	GA/ANN
Feeder- Solar Penetration	MAPE [%], RMSE [MW], R^2	MAPE [%], RMSE [MW], R^2
Valley center – 23.8%	5.06 0.10 0.96	5.24 0.10 0.96
Avocado – 13.3%	2.54 0.19 0.97	2.83 0.17 0.97
Creelman – 9.2%	2.39 0.05 0.95	3.34 0.05 0.95
Cabrillo – 5.8%	2.24 0.11 0.98	1.92 0.09 0.99
Alpine – 2.4%	2.40 0.07 0.98	2.76 0.06 0.99

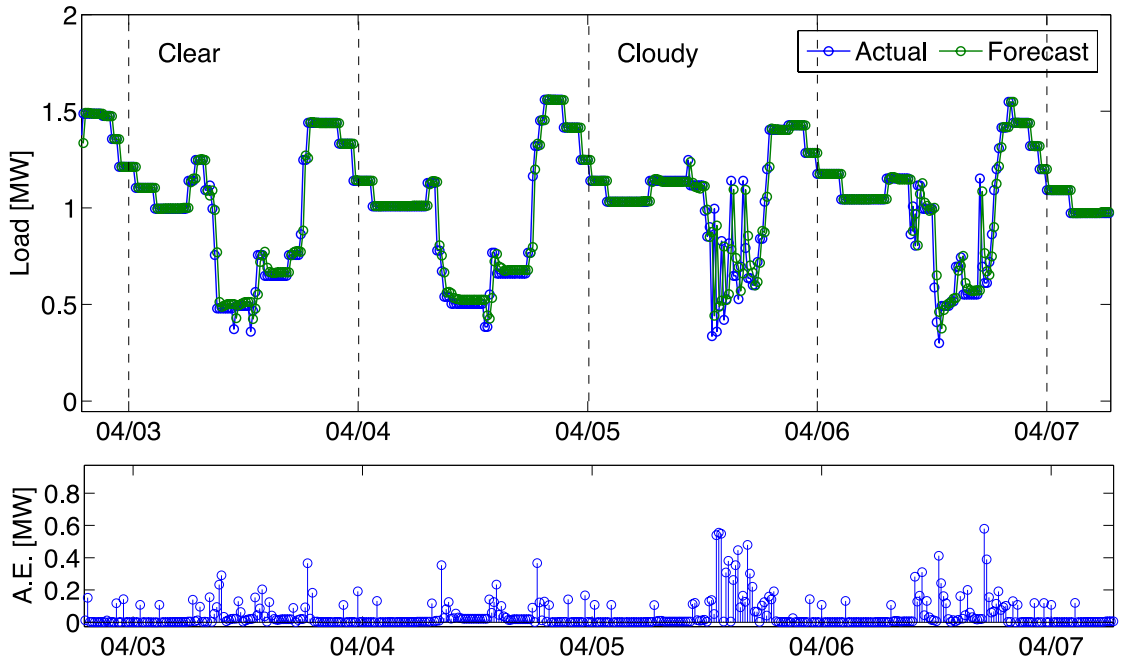

Figure 2: Load forecast results for the Valley Center feeder for 15 min forecast horizon using SVR. The first two days are clear days whereas the next two days depict cloudy days. The absolute error increases for cloudy days especially during solar production time. The above results are in PDT time. The vertical dashed lines indicate midnight.

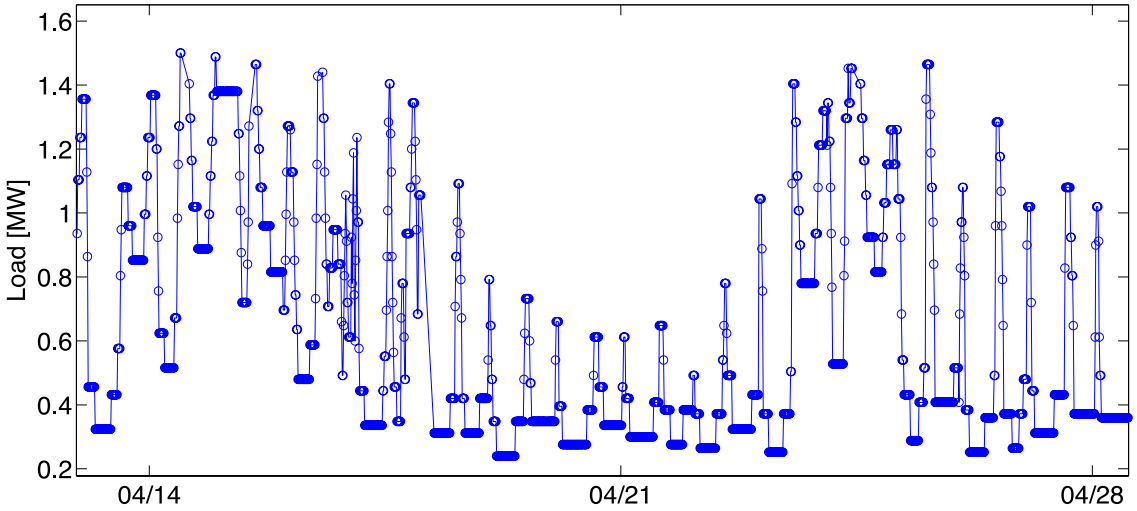

Figure 3: Load profile for the Creelman feeder for 04-14-2013 to 04-28-2013. It can be seen that the load profile for this feeder has high variability with sudden increase in load demand with no correlation to the past values. These spikes in load demand makes load forecasting very challenging with no information available about the sudden increase.

Furthermore, to analyze if forecasts can be refined further by learning from the past errors, autocorrelation between the forecast residuals was analyzed. Figure 7 shows the autocorrelation in the load forecasting errors for the 15 min forecast horizon. The autocorrelation is almost white

i.e., after 0 lag the errors are independent of each-other. It suggests that the model has captured all the information available in time-series and no information is left in the forecast residuals that could be leveraged to improve the forecast. Thus, this is the best forecast, which can be produced using such machine learning models with no exogenous input.

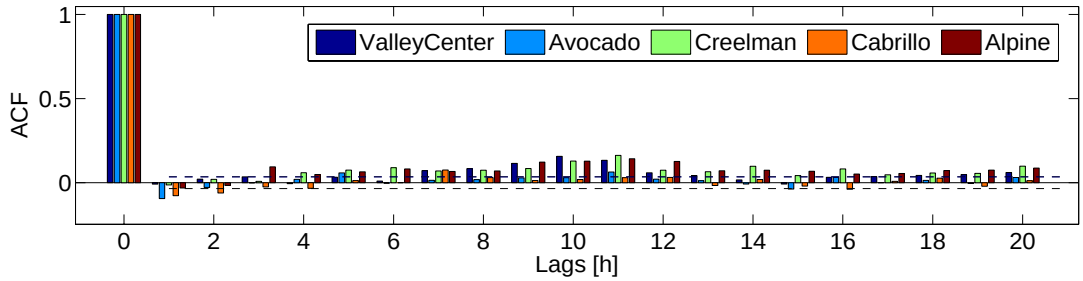

Figure 7: The autocorrelation for the residuals of the forecast for 15 min forecast horizon is almost white which validates the model identification and indicates all the information from the time-series has been captured by the model.

Solar variability and solar penetration

Next, the impact of solar variability and the level of solar penetration on intra-hour load forecast capabilities was studied using the synthetic high-resolution satellite-derived solar data GHI from Task 2. Ramps, $R(t)$ at time (t) for 30 min time-scale were defined as: $R(t) = GHI(t) - GHI\ (t - 1)$. Since the resolution for satellite-derived data was 30 min, separate forecasts were produced for the 30 min forecast horizon for 30 min averaged load time-series. Hence, the MAPE for these forecasts is lower than the forecast issued at 30 min forecast horizon for 15 min time-averaged values as many ramps at 15 min timescale average out at 30 min time-scale. It was observed that the load forecast error is independent of solar variability for feeders with low solar penetration. As solar penetration increases, the load forecast error increases and is a linear function of solar variability. As shown in Figure 8, the error increases linearly with increasing solar variability for the Valley Center feeder load which has highest solar penetration. In contrast to that, the error for Alpine (the feeder with lowest solar penetration) is uniform and stays the same irrespective of the solar variability.

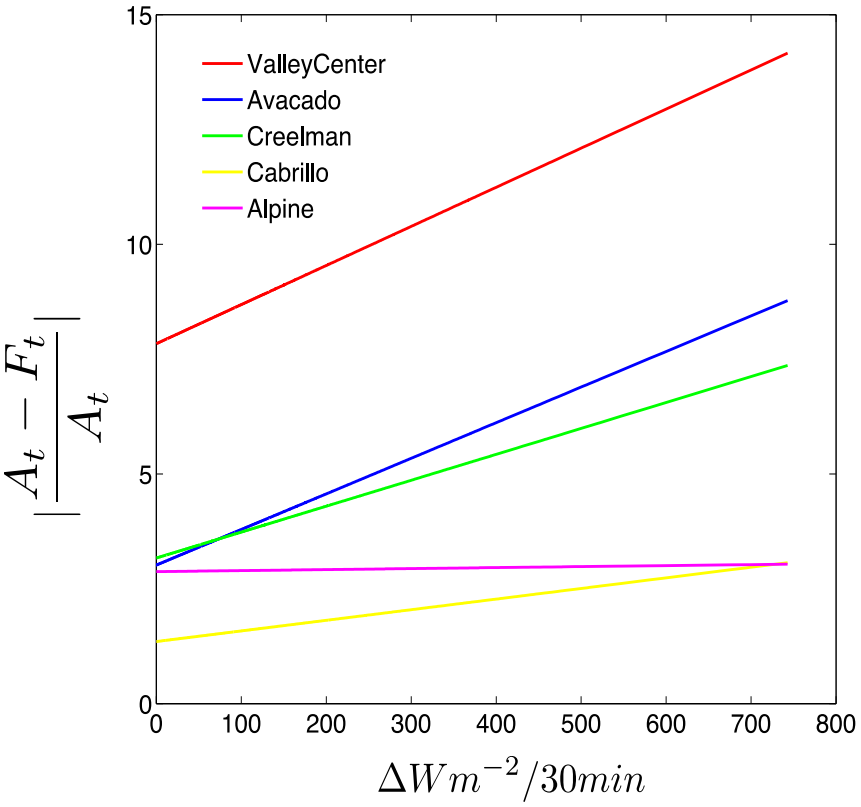

Figure 8: Normalized absolute forecast error for the 30 min forecast horizon using SVR versus solar variability for the five feeders. A_t is the actual value of the load, F_t is the forecast value for time t and Δ represents the change in Global Horizontal Irradiance at every 30 min time step. The error is independent of the solar variability for feeders with lowest penetration (e.g. Alpine). With increasing solar penetration, the load forecast error increases and appears to be a linear function of solar variability.

Days-ahead forecasts

Intra-day forecasts were implemented for 1 – 5 day ahead forecast horizon with hourly resolution. The MAPE for each feeder for all the forecast horizons is shown in Figure 4. Detailed error statistics are provided in appendix (Table 4-Table 8). These forecasts are based on SVR and kNN model. The input vectors are the lagged values for the same hour from the past seven days. As shown in Figure 4, the MAPE for all feeders increases with increasing forecast horizon, eventually approaching a constant value for instance after the 4 days forecast horizon, the MAPE stays constant. Since the forecast for more than four days-ahead is giving information just about the shape of load profile like a polynomial fit, the error remains constant. Contrary to the expected result that the feeder with the highest penetration would contain the highest MAPE, the Creelman feeder has the highest MAPE irrespective of the forecast horizon. This can be again attributed to its load profile as discussed above and shown in Figure 3.

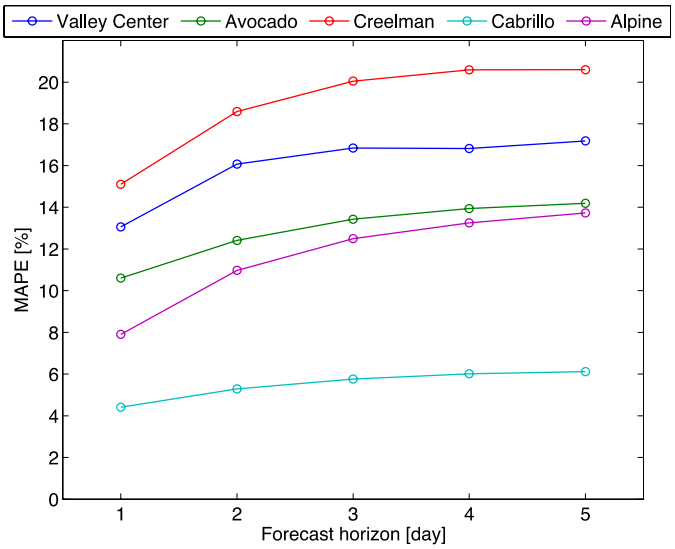

Figure 4: Mean Absolute Percentage Error for all the feeders using SVR.

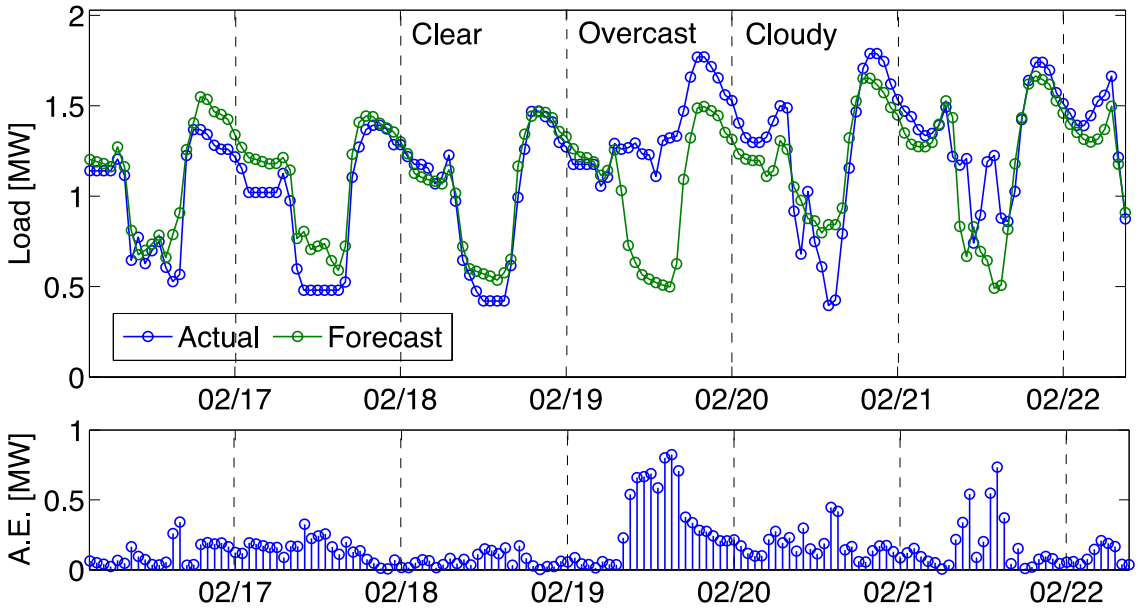

Figure 5: Load forecast results for the Valley Center feeder for the 24h forecast horizon (top) and Absolute Error (AE) at the same times (bottom) for the year 2013 using SVR. The AE increases during daytime due to varying solar generation. The timestamps are in PDT.

The above results were obtained without exogenous input. This demonstrates the challenging time-series characteristics for the day-ahead forecast horizon, in part due to ramps introduced by varying solar power generation. Sample intra-day forecasts are shown in Figure 5. To correct for these errors, an ensemble re-forecast method [5] was applied using day-ahead Global Horizontal Irradiance (GHI) predictions obtained from a Numerical Weather Prediction model. National Oceanic and Atmospheric Administration NAM forecasts were available only for the day-ahead

(36 hour) forecast horizon. Due to data availability, the NAM forecasts were utilized to enhance the results for the day-ahead forecast horizon only.

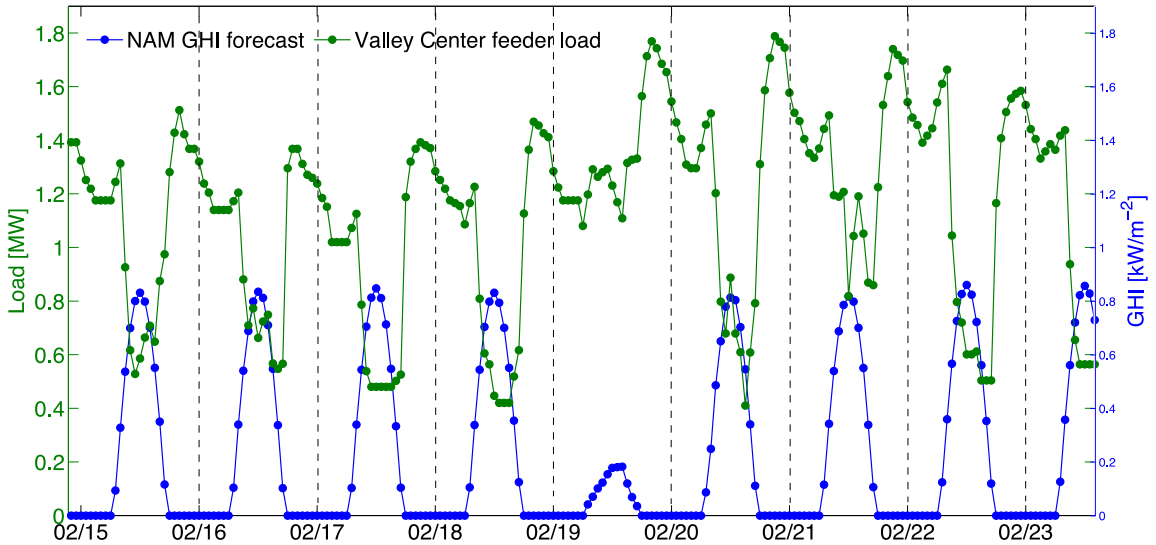

Figure 6: Time synchronized NWP GHI and feeder load from Valley Center for 02-15-2013 to 02-23-2013 PDT. The NWP model is able to predict the overcast day correctly. No ramps or variability in solar irradiance were predicted for partly cloudy conditions on the other days.

The improvement achieved using ensemble re-forecast methods was marginal. It was found that improvements in the forecast are limited by the accuracy of the solar forecasts. Figure 6 shows time synchronized feeder load data and the GHI forecasts issued at 00 UTC for the next day. It can be seen that for the same GHI forecast, the feeder has variable load profile with different ramps (02-20-2013 and 02-22-2013), which were likely caused by solar variability. Thus, even after using the NWP forecasts as input, it is very challenging to forecast the daytime ramps in the net load. For now, effects introduced by the spatial extend of the area covered by the feeder were assumed to be negligible. Therefore, the point forecasts for the feeder location were used in the model, which seems to be insufficient to capture the spatial variability in the net load. In future studies, the NWP forecasts for the locations with solar generation weighted based on the power capacity of the solar plants is recommended to be added as an input for the net load forecasting models.

Conclusions

Net-load forecast for intra-hour and multiple day-ahead forecasts were successfully implemented for the five feeders in SDG&E operating region. For the 15 min forecast horizon, both SVR and ANN-GA forecast models are recommended whereas for 30 min forecast horizon ANN-GA based models outperformed the SVR models with a MAPE error ranging from 7.07% for the feeder with highest penetration (Valley Center) to 3.78% for the feeder with lowest penetration (Alpine) except for Creelman feeder. Intra-hour forecast error for the 30 min forecast horizon increases linearly as a function of solar variability for high (>5%) solar penetration feeders. Furthermore, NWP day-ahead GHI forecasts were added as an exogenous variable, but the net load forecast improvements were marginal. Since NWP day-ahead forecasts in coastal California are generally inaccurate and most net load forecast error was driven by short-term solar variability that was unresolved in the NWP. Therefore, it is suggested to use forecasts from all the location with solar generation as an input and ensemble them based on the capacity of the solar generation plant.

References

[1] H.T.C. Pedro, C.F.M. Coimbra, Assessment of forecasting techniques for solar power production with no exogenous inputs, Solar Energy 86 (7) (2012) 2017 – 2028.
[2] V. Vapnik, S. E. Golowich, A. Smola, Support vector method for function approximation, regression estimation, and signal processing, Advances in neural information processing systems (1997) 281–287.
[3] B.J. Chen, M.W. Chang, C.J. Lin, Load forecasting using support vector machines: a study on EUNITE competition 2001, IEEE Transactions on Power Systems 19 (4) (2004) 1821–1830.
[4] A. Kaur, H.T.C. Pedro, C.F.M. Coimbra, Impact of onsite solar generation on system load demand forecast, Energy Conversion and Management 75 (0) (2013) 701 – 709.
[5] A. Kaur, H.T.C Pedro, C.F.M. Coimbra, Ensemble re-forecasting methods for enhanced power load prediction, Energy Conversion and Management 80 (0) (2014) 582 – 590.

Appendix

Table 3: Net load forecasting results for 30 min forecast horizon for the five feeders.

Models	SVR	GA/ANN
Feeders- Solar Penetration	MAPE [%], RMSE [MW], R^2	MAPE [%], RMSE [MW], R^2
Valley center – 23.8%	8.45 0.14 0.93	7.07 0.12 0.95
Avocado – 13.3%	4.63 0.24 0.94	3.94 0.21 0.96
Creelman – 9.2%	4.63 0.07 0.88	6.34 0.07 0.88
Cabrillo – 5.8%	3.23 0.16 0.97	2.40 0.12 0.98
Alpine – 2.4%	4.83 0.09 0.97	3.78 0.07 0.98

Table 4: Net load forecasting results for 1 day forecast horizon for the five feeders.

Models	SVR	kNN
Feeders- Solar Penetration	MAPE [%], RMSE [MW], R^2	MAPE [%], RMSE [MW], R^2
Valley center – 23.8%	13.06 0.22 0.81	14.49 0.22 0.82
Avocado – 13.3%	10.60 0.39 0.71	11.08 0.56 0.70
Creelman – 9.2%	15.10 0.13 0.57	17.23 0.13 0.53
Cabrillo – 5.8%	4.41 0.26 0.91	4.60 0.25 0.93
Alpine – 2.4%	7.90 0.19 0.87	9.06 0.21 0.83

Table 5: Net load forecasting results for 2 days forecast horizon for the five feeders using SVR.

Feeders- Solar Penetration	MAPE [%], RMSE [MW], R^2
Valley center – 23.8%	16.07 0.28 0.72

Avocado – 13.3%	12.41 0.64 0.60
Creelman – 9.2%	18.59 0.16 0.36
Cabrillo – 5.8%	5.29 0.29 0.90
Alpine – 2.4%	10.97 0.27 0.72

Table 6: Net load forecasting results for 3 days forecast horizon for the five feeders using SVR.

Feeders- Solar Penetration	MAPE [%], RMSE [MW], R^2
Valley center – 23.8%	16.84 0.31 0.64
Avocado – 13.3%	13.43 0.69 0.54
Creelman – 9.2%	20.05 0.17 0.27
Cabrillo – 5.8%	5.76 0.31 0.88
Alpine – 2.4%	12.49 0.32 0.62

Table 7: Net load forecasting results for 4 days forecast horizon for the five feeders using SVR.

Feeders- Solar Penetration	MAPE [%], RMSE [MW], R^2
Valley center – 23.8%	16.82 0.32 0.62
Avocado – 13.3%	13.94 0.71 0.52
Creelman – 9.2%	20.59 0.17 0.22
Cabrillo – 5.8%	6.01 0.32 0.87
Alpine – 2.4%	13.25 0.34 0.57

Table 8: Net load forecasting results for 5 days forecast horizon for the five feeders using SVR.

Feeders- Solar Penetration	MAPE [%], RMSE [MW], R^2
Valley center – 23.8%	17.18 0.33 0.60
Avocado – 13.3%	14.19 0.71 0.51
Creelman – 9.2%	20.60 0.17 0.22
Cabrillo – 5.8%	6.12 0.33 0.87
Alpine – 2.4%	13.73 0.35 0.54